

## Social media anxiety detection using prompt engineering techniques

A.L. Senanayake<sup>1</sup>, and S.R. Liyanage<sup>1\*</sup>

<sup>1</sup>Faculty of Computing & Technology, University of Kelaniya, Sri Lanka

Anxiety is the most common mental disorder, yet it receives less attention compared to other mental health issues. The exponential growth of social media has led people to share their feelings, moods, and tension through social media content, making it a good source for identifying the mental status of individuals. Analyzing social media texts provides valuable insights for screening mental issues such as anxiety. According to the literature, there are multiple approaches for the detection of anxiety via text classification like emotion analysis, sentiment analysis, topic modeling, language analysis, and deep learning-based approaches like bi-LSTM. Modern advancements in natural language processing have obtained state-of-the-art (SOTA) performance using transformer-based pre-trained language models (PLM) like PsychBERT, DisorBERT, and MentalBERT across various domains including digital mental health. These models mainly focus on depression, suicide, and stress. Fine-tuning these models for anxiety classification requires a substantial amount of computational resources and large annotated corpora. The capability to fine-tune models for specific tasks while keeping most of the PLM parameters frozen has led to significant success in parameter-efficient prompt engineering. Our approach introduces a novel framework that enhances domain-specific knowledge of PLMs for in-context learning of the detection of anxiety using crafted soft prompt templates. The data was pre-processed with PLM libraries, combined with soft templates to generate a prompt dataset, and the resulting embeddings were then used for tuning the prompt. For implementing the prompt engineering logic, a well-balanced, anxiety-annotated Reddit dataset which include 356,242 anxiety posts and 354,631 non-anxiety posts keeping the proportions of 70%, 15%, and 15% for training, validation, and testing. Our method surpasses MentalBERT SOTA performance in anxiety classification, achieving 80.5% F1 score and 81.8% accuracy, with maximum margins of 4.3% and 4.6%, respectively. Evaluation results show that the proposed method improves the ability of PLMs to classify anxiety by using prompt engineering techniques.

**Keywords:** Anxiety detection, digital mental health, natural language processing, social media

---

\*sidath@kln.ac.lk  
ORCID ID: <https://orcid.org/0000-0003-0778-0874>